

二、项目信息

注：请勿泄露个人信息，原则上不得出现任何与姓名、单位名称相关的信息，包括申请人姓名、设站单位名称、合作导师姓名等。可参考使用“本人”“本单位”“本实验室”等表述。

项目名称	基于鲁棒视觉特征的复杂场景密集预测方法研究		
项目来源	自选 <input checked="" type="checkbox"/> 合作导师项目 <input type="checkbox"/> 自主获得的国家或省部级课题 <input type="checkbox"/> 其他 <input type="checkbox"/>		
项目所属一级学科	计算机科学与技术	项目所属二级学科	计算机应用技术
交叉一级学科		交叉二级学科	

1. 选题依据（国内外研究现状及选题价值，限 1000 字）

自动驾驶系统通过先进的通讯、控制和计算机技术，在无人为干预条件下实现对车辆的实时连续控制，旨在为人们提供一种安全且顺畅的出行方式和物流解决方案。近几年，深度学习的兴起促进了新一轮人工智能的快速发展，自动驾驶不断与人工智能技术深度融合，正在重塑汽车及相关智能产业的生态体系。图像密集预测（主要包括单目深度估计^[1]及场景语义分割^[2]）是人工智能应用在自动驾驶领域的关键技术之一，可以提供丰富的视觉信息用于辅助自动驾驶系统进行决策。因此，实现高精度的密集预测是保障该技术在自动驾驶领域产业化落地的重要条件，具有重要的研究意义和应用价值。

凭借极强的学习能力，深度学习方法^[3,4]已经成功地应用于自动驾驶场景中的密集预测任务。在自动驾驶的复杂场景中，车载成像系统极易受到成像因素（例如分辨率、噪声及模糊等）和环境因素（例如光照和天气条件变化等）影响，造成获取图像的质量严重退化。对退化图像进行密集预测时，多数深度学习方法训练的模型很难提供精准的预测结果，无法为驾驶系统提供准确的辅助信息，严重影响密集预测模型的实际应用价值。因此，对于在安全性方面有严格要求的自动驾驶系统，迫切地需要深度学习模型具备良好的鲁棒性，保障模型在复杂场景下始终可以提供准确可靠的密集预测结果。

自动驾驶对深度学习模型的鲁棒性^[5,6]提出了更高的要求，天气环境变化是影响模型性能的关键因素之一，很多方法采用建模天气条件^[7,8,9]，或者构建关于恶劣天气的虚拟数据集^[10,11]和真实数据集^[12,13]等方式，提高深度学习模型对恶劣天气中退化图像的鲁棒性。为了解决雨天目标检测模型的脆弱性问题，Volk等人^[14]引入人工合成的降雨图像训练模型，提升了模型在雨天场景中的目标检测精度。Ahmed等人^[15]提出了一个基于先验增强的深度卷积神经网络，实现了恶劣天气下高精度的交通标志检测。Liu等人^[16]通过弱监督参数估计网络估计可辨识图像处理模块的超参数，可以显著提升不同恶劣天气条件下的目标检测精度。与目标检测相比，密集预测任务提供了细粒度的视觉信息，可以辅助自动驾驶系统更好地感知周围环境变化。Tremblay等人^[17]利用物理粒子模拟器构建虚拟雨天数据集，通过该数据集微调参数后，模型在深度估计和语义分割任务上的性能得

到显著提高。Kamann等人^[18, 19]探索不同网络结构在提高语义分割模型鲁棒性方面发挥的作用。Pan等人^[20]通过图像复原方法恢复模糊图像的视觉质量，以此方式提高模糊图像的深度估计精度。Dai等人^[21]利用域适应方法提高模型对雾天图像的语义理解能力。Yan等人^[22]搭建了一个包含光流估计网络和域对齐网络的一体化框架，显著提高了模型在浓雾图像上的光流估计精度。在自动驾驶场景下，Zhou等人^[23]提出了一种用于自动评估语义分割模型鲁棒性的新方法。Klingner等人^[24]在多任务学习框架下利用深度估计辅助语义分割任务，以此方式提升模型对噪声攻击的鲁棒性。Choi等人^[25]提出了一种实例选择性损失函数，可以保持域不变特征的同时移除导致域差异的风格信息，显著地提升了语义分割模型对不同气候变化的鲁棒性。Wu等人^[26]提出了一种无监督的域适应方法，通过对抗训练将白天数据集训练的模型迁移到夜间场景下，实现了夜间高精度的语义分割。

关于如何提高深度学习模型在复杂场景中鲁棒性的课题吸引了众多研究者的关注，随着数据集的不断丰富以及评价体系的逐渐完善，很多鲁棒的深度密集预测模型相继提出，但该领域仍然存在很多问题亟待解决：（1）泛化性较差。通过域适应方法可以将模型迁移到特定退化图像，虽然对特定类型退化的鲁棒性得到提高，但模型仍然无法在其他类型的退化图像上获得可靠的密集预测精度；（2）整体性能降低。通过数据扩充方式可以提高模型对特定退化图像的鲁棒性，但同时也影响了模型在不同类型退化图像上的整体性能，甚至一定程度上降低了模型在未退化图像上的性能；（3）鲁棒性偏弱。很多方法尝试提高模型应对不同退化图像的鲁棒性，但由于同时抵抗图像多种退化模式的能力较弱，模型在退化图像上的性能距离自动驾驶的应用需求仍然相差甚远。

综上所述，提高深度学习模型在复杂场景中的鲁棒性具有极强的应用需求。目前关于这一课题的研究仍处于初期探索阶段，要实现最终成熟的应用技术仍然需要进一步探索。鉴于此，本项目将深入研究基于鲁棒视觉特征的复杂场景密集预测方法，以“映射空间”为出发点，以“网络架构”为拓展点，以“模型优化”为突破点，发挥正交矩和深度学习的协同优势，提高密集预测技术在自动驾驶应用中的安全性和可靠性。由此可见，本项目研究成果对推动人工智能在自动驾驶领域的产业化落地具有重要作用。

参考文献

- [1] S. Bhat, I. Alhashim, P. Wonka. AdaBins: Depth estimation using adaptive bins. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 4009-4018.
- [2] 秦飞巍, 沈希乐, 彭勇, 邵艳利, 袁文强, 计忠平, 白静. 无人驾驶中的场景实时语义分割方法. 计算机辅助设计与图形学学报, 2021, 33(07): 1026-1037.
- [3] S. Vandenhende, S. Georgoulis, W. Gansbeke, M. Proesmas, D. Dai, L. Gool. Multi-task learning for dense prediction tasks: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2022, 44(7): 3614-3633.
- [4] D. Bruggemann, M. Kanakis, A. Obukhov, S. Georgoulis, L. Gool. Exploring relational context for multi-task dense prediction. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2021: 15849-15858.
- [5] 纪守领, 杜天宇, 邓水光, 程鹏, 时杰, 杨珉, 李博. 深度学习模型鲁棒性研究综述. 计算机学报, 2022, 45(01): 190-206.
- [6] G. Chen, P. Peng, L. Ma, J. Li, L. Du, Y. Tian. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2021: 448-457.

- [7] 解宇虹, 谢源, 陈亮, 李翠华, 曲延云. 真实有雾场景下的目标检测. 计算机辅助设计与图形学学报, 2021, 33(05): 733-745.
- [8] 高君宇, 杨小汕, 张天柱, 徐常胜. 基于深度学习的鲁棒性视觉跟踪方法. 计算机学报, 2016, 39(07): 1419-1434.
- [9] S. Huang, T. Le, D. Jaw. DSNet: Joint semantic learning for object detection in inclement weather conditions. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021, 43(8): 2623-2633.
- [10] C. Sakaridis, D. Dai, L. Gool. Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision (IJCV), 2018, 126(9): 973-992.
- [11] M. Johnson-Roberson, C. Barto, R. Mehta, S. Sridhar, K. Rosaen, R. Vasudevan. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 2017: 746-753.
- [12] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, T. Darrell. BDD100K: A diverse driving dataset for heterogeneous multitask learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 2633-2642.
- [13] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, R. Yang. The ApolloScape open dataset for autonomous driving and its application. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020, 42(10): 2702-2719.
- [14] G. Volk, S. Muller, A. Bernuth, D. Hospach, O. Bringmann. Towards robust CNN-based object detection through augmentation with synthetic rain variations. In: Proceedings of International Conference on Intelligent Transportation Systems (ITSC), 2019: 285-292.
- [15] S. Ahmed, U. Kamal. M. Hasan. DFR-TSD: A deep learning based framework for robust traffic sign detection under challenging. IEEE Transactions on Intelligent Transportation Systems (TITS), 2022, 23(6): 5150-5162.
- [16] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, L. Zhang. Image-adaptive YOLO for object detection in adverse weather conditions. In: Proceedings of AAAI Conference on Artificial Intelligence (AAAI), 2022: 1792-1800.
- [17] M. Tremblay, S. Halder, R. Charette, J. Lalonde. Rain rendering for evaluating and improving robustness to bad weather. International Journal of Computer Vision (IJCV), 2021, 129(2): 341-360.
- [18] C. Kamann, C. Rother. Benchmarking the robustness of semantic segmentation models. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 8825-8835.
- [19] C. Kamann, C. Rother. Benchmarking the robustness of semantic segmentation models with respect to common corruptions. International Journal of Computer Vision (IJCV), 2021, 129(2): 462-483.
- [20] L. Pan, S. Chowdhury, R. Hartley, M. Liu, H. Zhang, H. Li. Dual pixel exploration: Simultaneous depth estimation and image restoration. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 4340-4349.
- [21] D. Dai, C. Sakaridis, S. Hecker, L. Gool. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. International Journal of Computer Vision (IJCV), 2020, 128(5): 1182-1204.
- [22] W. Yan, A. Sharma, R. Tan. Optical flow in dense foggy scenes using semi-supervised learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 13256-13265.
- [23] W. Zhou, J. Berrio, S. Worrall, E. Nebot. Automated evaluation of semantic segmentation robustness for autonomous driving. IEEE Transactions on Intelligent Transportation Systems (TITS), 2020, 21(5): 1951-1963.
- [24] M. Klingner, A. Bar, T. Fingscheidt. Improved noise and attack robustness for semantic segmentation by using multi-task training with self-supervised depth estimation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020: 1299-1309.
- [25] S. Choi, S. Jung, H. Yun, J. Kim, S. Kim, J. Choo. RobustNet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In: Proceedings of IEEE Conference on

Computer Vision and Pattern Recognition (CVPR), 2021: 11580-11590.

[26] X. Wu, Z. Wu, H. Guo, L. Ju, S. Wang. DANNet: A one-stage domain adaptation network for unsupervised nighttime semantic segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 15769-15778.

2. 研究内容（研究对象，拟解决的关键科学问题，研究目标，限 2000 字）

2.1 研究目标

密集预测是人工智能应用在自动驾驶领域的关键技术之一，也是本项目的核心研究课题。密集预测提供了关于周围环境丰富的细粒度视觉信息，可以辅助自动驾驶系统更好地感知周围环境变化，及时准确地对车辆行驶进行控制。近年，基于卷积神经网络的深度学习方法逐渐兴起，促进了密集预测技术的快速发展。但是，由于受到恶劣气候或者成像因素干扰，自动驾驶场景获取图像的质量退化变得极其常见。当前基于深度学习的密集预测方法很难在质量退化的图像上进行精准预测，难以满足自动驾驶系统对高安全性和可靠性的严格要求，降低了密集预测技术的实际应用价值。针对上述问题，本项目将探索如何提高深度学习模型的鲁棒性，保障模型在复杂场景的不同类型退化图像上均可实现高精度的密集预测，提高密集预测技术在自动驾驶应用场景中的安全性和可靠性，为推动以密集预测为代表的人工智能技术产业化落地打下坚实的基础。

2.2 研究对象与研究内容

本项目旨在研究基于鲁棒视觉特征的复杂场景密集预测方法，考虑正交矩的良好稳定性和深度学习极强的学习能力，将分别以“映射空间”为出发点，以“网络架构”为拓展点，以“模型优化”为突破点，发挥正交矩和深度学习的协同优势，构建复杂场景下鲁棒密集预测的完备体系。图1为本项目的研究框架，下面将介绍具体的研究内容。

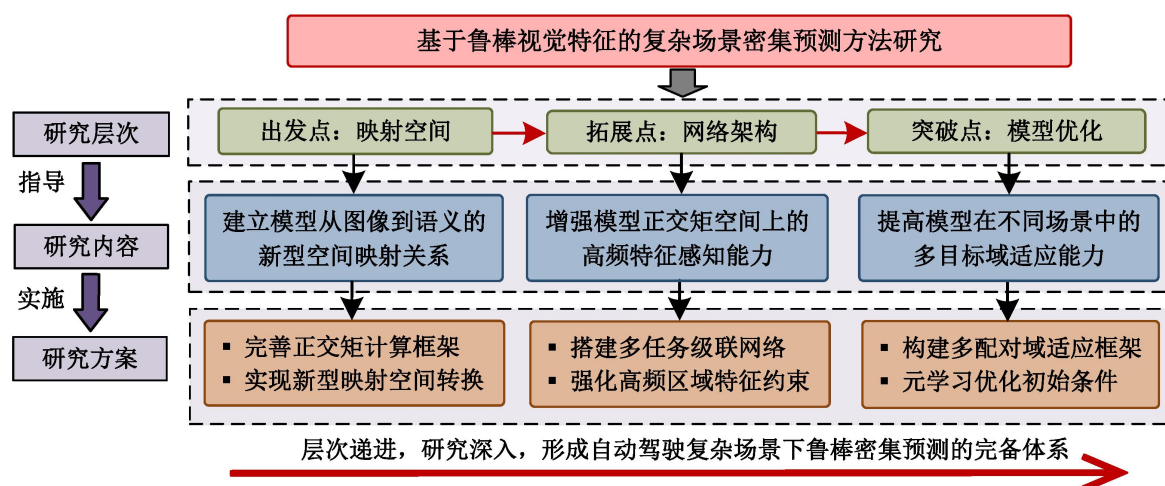


图1 本项目的总体研究框架

(1) **映射空间**: 建立深度学习模型稳定的空间映射关系。受到成像因素和环境因素影响，自动驾驶复杂场景中获取图像的质量极易发生退化。即使训练阶段通过模拟退化图像对数据集进行扩充，也无法保证模型很好地泛化到其他类型的退化图像。造成上述问题的主要原因是，退化图像的像素在干扰下通常会发生剧烈变化，破坏了模型学习出

映射关系的稳定性，造成模型无法预测出每个像素准确的语义信息。

针对上述问题，本项目拟从映射空间角度提出一种简单且有效的策略，利用稳定的空间替换图像像素空间，实现连接语义空间更加稳定的映射关系，以期挖掘鲁棒的视觉特征，为不同类型退化图像提供高精度密集预测结果。围绕映射空间，本项目将首先改良正交矩计算方法，在降低误差的同时提高计算速度，然后探索不同阶数和重复度正交矩的稳定性并选择最优组合计算正交矩，最后利用稳定正交矩构建新型映射空间，通过仿真实验验证不同正交矩与深度神经网络在复杂场景下密集预测的鲁棒性。

(2) 网络架构：增强正交矩空间对高频特征的感知能力。图像的高频区域涵盖了图像的全局拓扑结构信息以及局部的边缘和纹理细节。对于密集预测任务，由于高频区域的语义分布不连续且变化较大，深度学习模型在这些区域做出错误预测的风险明显大于平滑的低频区域。当图像质量受到严重破坏时，与平滑的低频区域相比，图像高频区域的像素值会发生更加剧烈的变化。即使正交矩空间具有良好的鲁棒性，高频区域上正交矩空间的稳定性也会受到影响，降低了这些重要区域上密集预测的精度。

针对上述问题，本项目拟从网络架构设计入手进行深入研究，利用多任务学习策略搭建高频感知的深度卷积神经网络，通过挖掘频特征感知和密集预测任务间相互依赖关系，在正交矩空间的高频特征学习过程加入更强的约束，提高模型正交矩空间在敏感高频区域的鲁棒性。围绕网络架构，本项目首先搭建双任务级联框架，渐进式地实现高频感知和密集预测，然后通过互信息熵损失在正交矩的高频区域加入隐式约束，最后联合隐式和显式约束，提高模型高频区域预测结果和标注语义的高阶一致性。

(3) 模型优化：提高复杂场景对多种退化域的泛化能力。对于深度学习模型，优化策略在提高泛化性方面发挥了重要作用。不同的气候条件和成像因素会产生不同类型退化图像，同类型的退化图像一般服从相似的数据分布。交叉域适应方法通过拉近特征空间分布可以将模型迁移到特定的目标退化域，但无法保证模型很好地泛化到训练阶段未接触类型的退化域，并未在本质上提升模型对所有类型退化图像的鲁棒性。

针对上述问题，本项目拟从模型优化角度入手进行深入研究，通过增强模型的泛化能力提高复杂场景下不同类型退化图像的密集预测精度。元学习是一种学习如何学习的策略，在不改变网络结构的情况下通过改变初始条件（即模型初始化参数）来提高模型的泛化性。围绕模型优化，本项目首先构建多种类型的退化域训练数据，然后搭建基于多域配对的域适应框架，最后采用元学习策略对模型初始参数进行优化，学习出每次迭代模型从何种初始条件可以更好地泛化到其他类型的目标退化域。

2.3 拟解决的关键科学问题

(1) 正交矩的快速精准计算。正交矩对图像退化具有良好的鲁棒性，但积分误差等因素影响了正交矩的数值稳定性。此外，实现映射空间的快速转换，需要低复杂度的正交矩计算。如何实现正交矩的快速稳定计算是本项目要解决的关键科学问题。

(2) 结合正交矩的网络搭建。选择不同阶数和重复度计算正交矩，将影响模型密集预测的速度和精度。此外，正交矩空间的高频感知能力与重要区域的鲁棒性密切相关。如何有效结合正交矩和深度神经网络是本项目要解决的另一个关键科学问题。

(3) 交叉域适应的元学习优化。域适应方法可以提高目标退化域的鲁棒性，但无法保

证模型很好地泛化到未知的退化域。元学习可以辅助模型学习出泛化未知域的先验知识，如何采用合理的元学习策略进行优化也是本项目要解决的关键科学问题。

3. 研究方案（限 2000 字）

本项目采用理论分析、算法设计、仿真实验相结合的研究方法，大量阅读相关研究文献，针对研究目标和拟解决的关键科学问题，结合已有的研究基础开展研究，将分别从映射空间，网络架构及模型优化三个角度入手，探索出如何通过正交矩和深度学习生成鲁棒的视觉特征，最终形成一套自动驾驶复杂场景下鲁棒密集预测的完备体系。

1. 建立模型从图像到语义间新型的空间映射关系：从图像像素空间到正交矩空间转换。

从映射空间转换角度出发，本项目将探索如何构建更加稳定的空间映射关系。图2展示本项目搭建的基于新型映射空间的密集预测方法，旨在通过发挥正交矩和深度学习的协同优势，将图像像素值替换为正交矩幅值，使深度学习模型由“图像到语义”映射转换到更加稳定的新型“正交矩到语义”空间映射。由于正交矩空间具有极强的抗攻击能力，通过正交矩建立的新型空间映射关系可以保障深度学习模型在训练和测试阶段更高的稳定性和一致性，使深度模型挖掘出可以进行准确密集预测的鲁棒视觉特征。

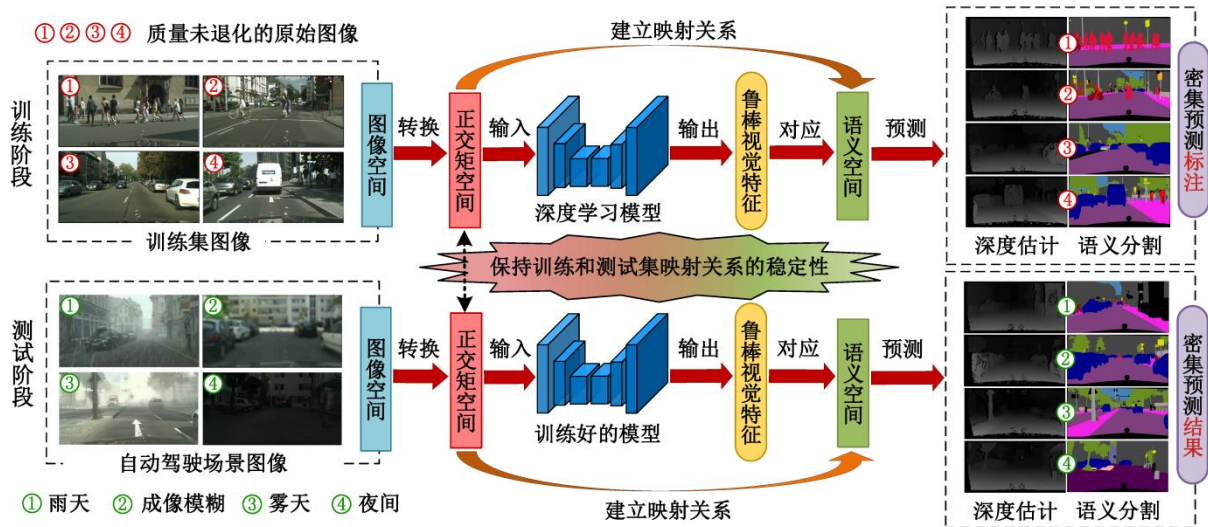


图2 基于正交矩空间映射的密集预测方法

正交矩空间不仅对噪声有极强的抵抗能力，还具有旋转、缩放和平移等几何不变性。但如何利用正交矩的鲁棒性优势，将其与深度学习进行有效结合，却始终未得到研究者的关注。正交矩是将图像投影到其核函数上得到的系数矩阵，正交矩的核函数由径向基函数和角向因子构成。对于极坐标图像函数 $f(r, \theta)$ 的正交矩 $Z_{m,n}$ ，其计算公式如下所示：

$$Z_{m,n} = \int_0^{2\pi} \int_0^1 f(r, \theta) R_{mn}(r) \exp(-jn\theta) r dr d\theta, \quad (1)$$

其中， m ， n 分别表示阶数和重复度， j 表示虚数单位， $R_{mn}(r)$ 表示径向基函数。围绕如何设计 $R_{mn}(r)$ ，出现了很多新型的正交矩，主要包括泽尼克矩（ZM）、伪泽尼克矩（PZM）、切比雪夫-傅里叶矩（CHFM）、指数-傅里叶矩（EFM）、极谐复指数变换（PCET）以及极谐傅里叶矩（PHFM）等。径向基函数的零点数量对应于径向基函数描述图像高频

分量的能力，通常情况零点数量越多，则描述能力越强。在表1和图3中深入剖析了几种典型正交矩的径向基函数以及对应的零点数量。

表1 不同正交矩径向基函数及对应的零点数量

正交矩	径向基函数 $R_{mn}(r)$	零点数量
ZM	$R_{mn}(r) = \sum_{s=0}^{(m- n)/2} \frac{(-1)^s (m-s)! r^{m-2s}}{s! ((m+ n)/2-s)! ((m- n)/2-s)!}$	$\frac{m- n }{2}$
PZM	$R_{mn}(r) = \sum_{s=0}^{m- n } \frac{(-1)^s (2m+1-s)! r^{m-s}}{s! (m+ n +1-s)! (m- n -s)!}$	$m- n $
CHFM	$R_m(r) = \sqrt{\frac{8}{\pi}} \left(\frac{1-r}{r} \right)^{\frac{1}{4}} \sum_{s=0}^{(m+2)/2} \frac{(-1)^s (m-s)! (4r-2)^{m-2s}}{s! (m-2s)!}$	m
EFM	$R_m(r) = \sqrt{\frac{1}{2\pi r}} \exp(j2m\pi r)$	$2m$
PCET	$R_m(r) = \sqrt{\frac{1}{\pi}} \exp(j2m\pi r^2)$	$2m$
PHFM	$R_m(r) = \begin{cases} 1/\sqrt{2}, m=0 \\ \sin[(m+1)\pi r^2], m \text{ 是奇数} \\ \cos[(m)\pi r^2], m \text{ 是偶数} \end{cases}$	$2m$

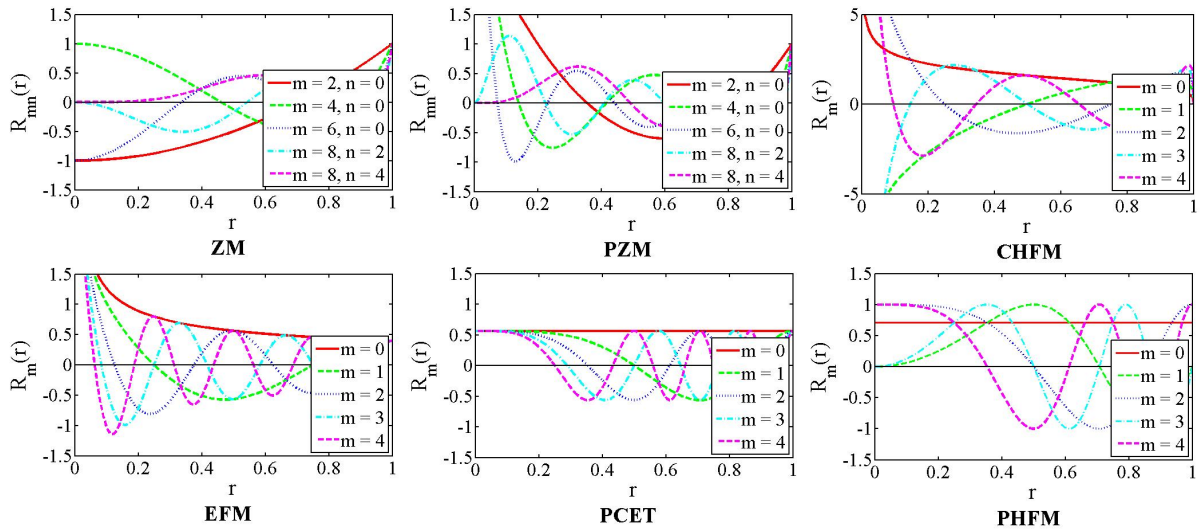


图3 不同正交矩的径向基函数分布

图4展示由图像空间到正交矩空间的转换过程：步骤1在输入图像上以每个像素为中心创建固定大小的窗口，步骤2将每个窗口中的图像内容由直角坐标系表示转换到极坐标系表示，步骤3选用表1中正交矩的径向基函数，按照公式1计算出正交矩幅值，步骤4利用稳定的矩幅值替换原始窗口的中心像素值，实现图像像素空间到正交矩空间转换。

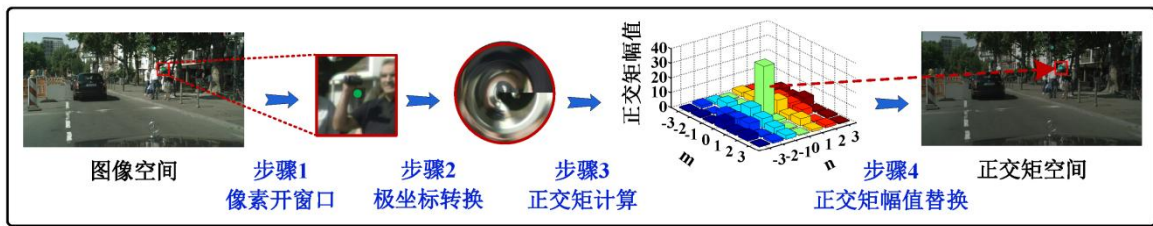


图4 正交矩空间转换流程

图5展示在不同类型退化时，图像局部窗口的矩幅值分布情况。可以看出，不同退化图像局部区域的正交矩幅值及幅值的分布极其稳定，证明了正交矩空间良好的鲁棒性。在训练阶段将数据集图像转化为矩空间，通过模型训练建立起正交矩到语义空间的映射关系。面对真实场景下不同类型的退化图像，将利用同样策略将测试图像转换到正交矩空间，然后利用训练好的模型进行密集预测。正交矩空间保证了训练和测试阶段映射关系的高度一致性，可以显著提高密集预测模型在复杂自动驾驶场景中的鲁棒性。

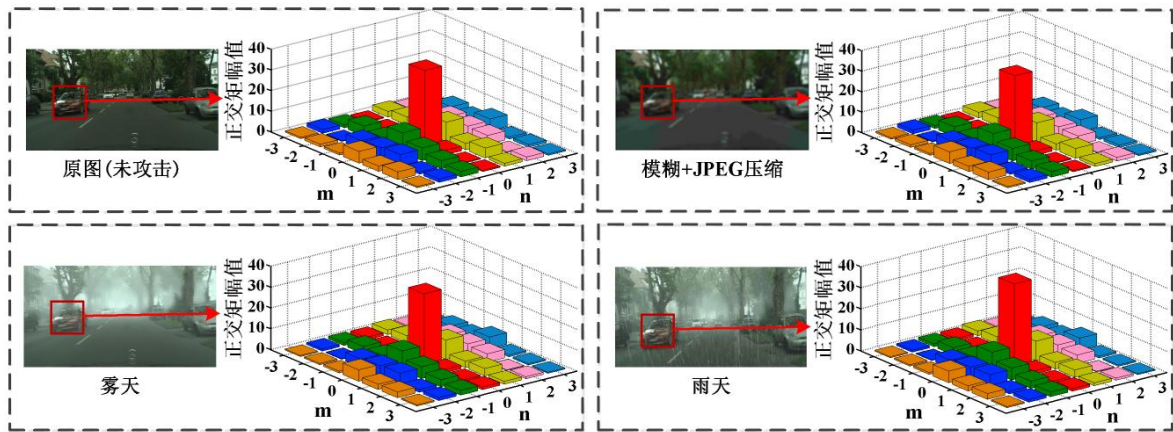


图5 不同退化图像局部窗口的稳定正交矩幅值分布

针对映射空间转换，本项目提出的方案简单且有效，但仍然存在一些问题需要解决：

- (1) 传统计算方法极易受到积分误差影响，如图3所示，当单位圆半径趋近0 ($r \rightarrow 0$) 时，径向基函数趋于无穷 ($R_{mn} \rightarrow \infty$)，将影响正交矩的描述能力并降低其数值稳定性。
- (2) 作为映射空间转换的预处理过程，需要正交矩更加快速的计算框架。本项目将进一步完善和丰富正交矩的理论体系与计算框架，探索如何设计更加快速、稳定的正交矩技术方法，为有效结合正交矩和深度学习模型进行鲁棒密集预测提供有力支撑。

2. 增强正交矩空间对高频特征感知能力：基于多任务的高频感知神经网络构建。

围绕网络架构设计，本项目将进一步探索如何更有效地结合正交矩和深度神经网络。图像的高频区域涵盖了图像的全局结构以及局部边缘和纹理，对于密集预测任务，图像高频区域附近的语义信息变化较大，造成模型极易在这些高频区域给出错误的预测结果。当图像质量受到破坏时，相对平滑的低频区域，高频区域的像素值也会发生更加剧烈的变化。即使正交矩具有极强的鲁棒性，高频区域正交矩的稳定性也会受到影响。为此，本项目将搭建高频感知的深度卷积神经网络，通过在正交矩空间的高频区域特征上加入更强的约束，确保模型在高频区域预测结果与真实语义内容的高阶一致性，以此方式进

一步有效融合正交矩和深度学习模型，提高复杂场景下退化图像的密集预测精度。

图6展示包含 T 个阶段的高频感知网络，首先将输入图像 x 转换到正交矩空间集合 Y ，其中红色和绿色六边形分别表示通过3阶和5阶正交矩计算出的两种正交矩空间。在第一阶段，首先采用通道拼接方式组合两种正交矩空间，并将其作为两个骨干网络的输入，然后通过两个网络分别预测出密集预测结果 P_1 和边缘检测结果 E_1 ，第一阶段中用于密集预测和边缘检测的骨干网络将分别选用高分辨率网络和双向级联网络。从第二阶段开始，每个阶段的输入包含前一阶段的预测结果和两种正交矩空间 Y ，每个阶段都采用相同的编码-解码结构，利用编码器挖掘出显式的引导信息（感知特征空间的高频区域），然后通过两个解码器分别预测当前阶段的密集预测结果和边缘检测结果。在第 T 阶段，网络通过多阶段的交互学习挖掘出鲁棒的视觉特征，借此预测出高精度边缘检测和密集预测结果。

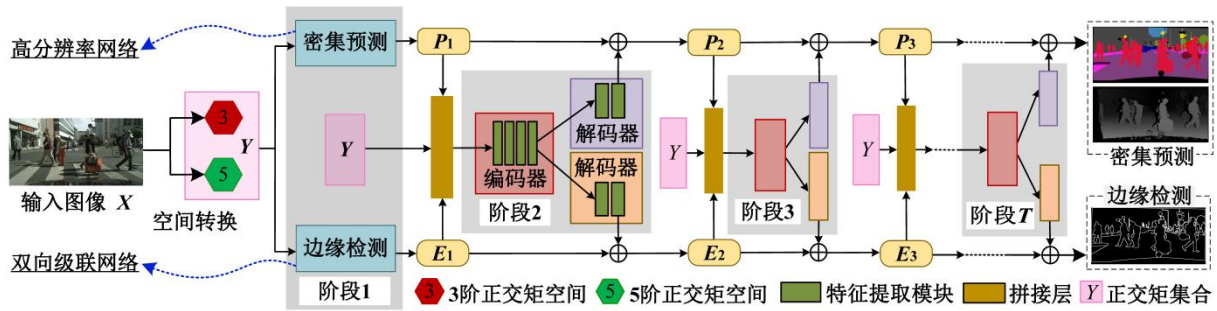


图6 基于多任务学习的高频感知网络

很多深度学习方法通过多任务学习策略提升密集预测精度，图7展示了两种比较典型的多任务学习策略以及本项目拟实施的多任务策略。第一种策略采用硬参数共享机制，通过辅助任务获取到显式的单向引导知识，然后利用这些知识优化密集预测结果。第二种策略采用软参数共享机制，从密集预测和辅助任务中挖掘出两种相互分离的显式引导信息，然后利用从其中一个任务挖掘出的引导信息优化另一个任务的预测结果。以上两种多任务学习策略，均属于简单地融合显式的引导信息。与之相比，本项目将进一步强化辅助任务的引导作用，通过辅助任务挖掘隐式的引导信息，然后利用这些信息构造一种新型的隐式损失函数，实现进一步约束高频区域上预测结果和人工标注的高阶一致性。

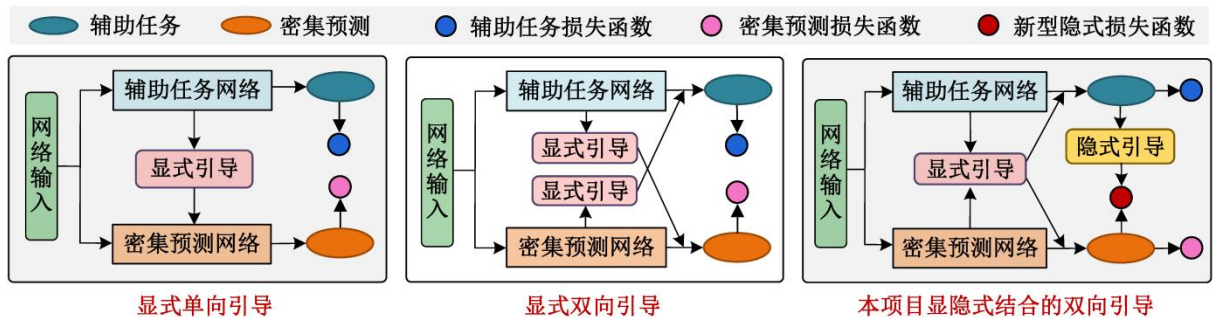


图7 多任务学习策略对比

密集预测任务通常使用交叉熵损失函数计算误差，忽略了像素之间存在的相关性。与之相比，交互信息熵损失函数充分考虑局部区域内像素的相关性，对于任意像素，定义以该像素为中心的局部区域，每个区域密集预测和人工标注由两个分布建模，通过最

小化两个分布互信息熵的下限可以最大化两个分布的相似性。复杂场景图像的高频区域呈现复杂的纹理变化，密集预测方法在这些高频区域普遍存在鲁棒性较差的问题。本项目拟提出一种边缘引导的新型隐式损失函数，其计算过程如图8所示，在多任务高频感知网络中，首先通过边缘检测结果上任意位置 E_0 ，分别在密集预测结果和人工标签上定位出对应位置 P_0 和 Y_0 ，接着创建不同大小的局部窗口，并计算窗口中两种阶数正交矩空间和对应标注区域的交互信息熵，最后通过最大化下界的方式保持高频区域的高阶一致性。

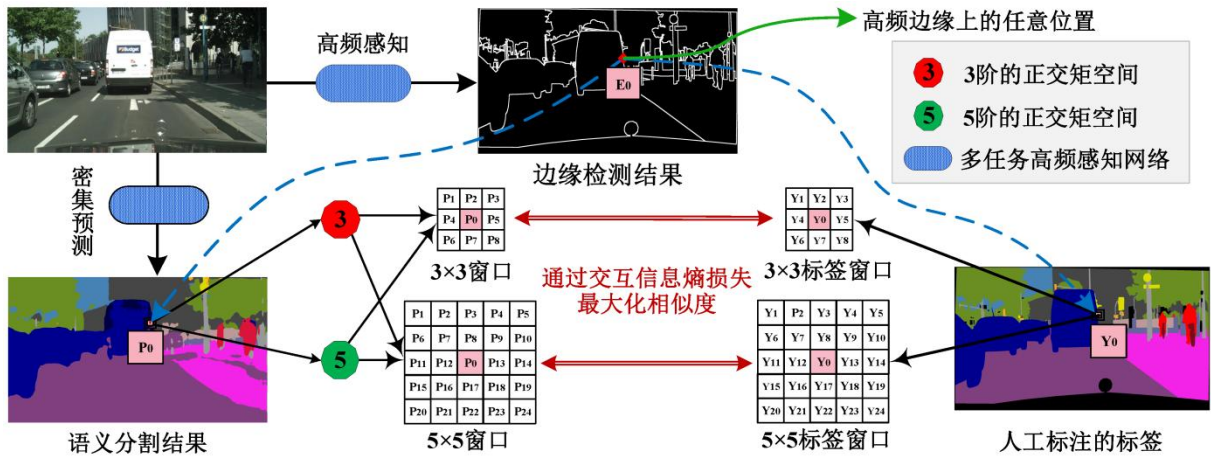


图8 基于高频边缘引导的隐式交互信息熵损失函数

针对网络架构设计，本项目拟搭建一种基于多任务的高频感知网络，使正交矩和深度学习模型得到更加有效的结合，利用多任务学习为密集预测提供显式和隐式指导，为高频区域的密集预测提供了更强的约束，进一步提高模型在重要高频区域密集预测的鲁棒性。围绕网络架构，项目仍然有一些科学问题需要进一步探索：（1）多任务框架结构设计方面，将探索如何增强模型挖掘高频特征的能力；（2）高频区域约束方面，将探索如何加强对正交矩空间高频区域的高阶约束。在映射空间转换基础上，本项目将进一步从网络结构方面入手，深入探索出正交矩提高密集预测模型鲁棒性的有效途径。

3. 提高模型不同场景下多目标域的泛化能力：正交矩空间的多域配对元学习优化。

围绕模型优化，本项目将进一步探索如何通过优化策略提高特征表示的鲁棒性和泛化性。对于深度学习方法，训练过程的优化策略在提高模型判别能力和鲁棒性方面的发挥了重要作用。这部分的研究目标是将密集预测模型更好地迁移到不同类型的退化图像上，通过增强模型的泛化能力提高各种复杂场景下退化图像的密集预测精度。同类型的退化图像的一般服从相似的数据分布，可以将同类型退化图像的正交矩空间作为特定的退化域。通过多种退化域构建出多退化域配对的域适应框架，然后采用元学习策略对模型参数进行优化，以此提升密集预测模型在多个退化域的鲁棒性。

元学习是一种学习如何学习的策略，可以在不改变网络结构的情况下，通过改变网络初始参数来提高模型的泛化性。本项目使用的训练集包含多种类型的退化图像，包括受到成像因素和天气因素影响的退化图像。在元学习优化时，首先将未受外界因素影响图像作为源域，并从训练集中任意选取两种退化域分别作为潜在域和目标域。然后从配

对“源域到潜在域”中学习元知识，将这些知识用于指导另一个配对“潜在域到目标域”的域适应过程。其中，元知识指模型从何种初始条件可以更好地迁移到目标退化域上。

令 D_S 表示源域， $D_L = \{D_L^0, D_L^1, \dots, D_L^{K-1}\}$ 和 $D_T = \{D_T^0, D_T^1, \dots, D_T^{K-1}\}$ 分别表示潜在退化域和目标退化域， K 表示退化域数量。在基于生成对抗的域适应方法中，模型由密集预测网络和域判别网络组成，通过无监督域损失函数 L_{uda} 来优化域适应模型，可以表示为 $L_{uda} = L_{pre} + L_{adv}$ 。其中， L_{pre} 表示密集预测网络的交叉熵损失， L_{adv} 表示域判别网络对抗损失， λ 表示用来平衡两个损失的权重。在密集预测和域判别网络之间加入梯度翻转层，可以使整个模型通过端到端的方式进行优化。如图9所示，在元学习优化过程中包含两个域适应过程，分别为源域到潜在域和潜在域到目标域。在元训练阶段，将元优化首先应用到源域到潜在域的域适应过程上，学习出代表模型初始条件的元知识，这些知识可以指导模型从何种初始条件出发可以更好地从带标注数据迁移到无标注数据上。在元测试阶段，学习出的元知识可以指导模型更好地从潜在域泛化到目标退化域。

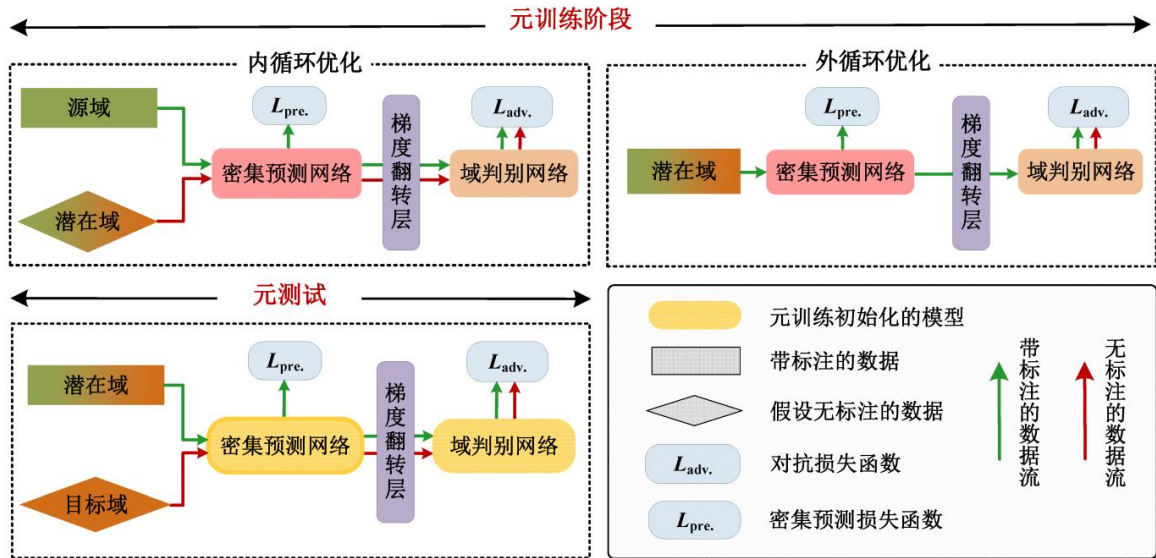


图9 多域配对框架的元学习优化

元训练阶段。将初始条件（即模型初始化参数）的元学习优化过程看作是一个包含内循环和外循环的双层优化问题，可以表示为如下形式：

$$\theta = \arg \min_{\theta} \overbrace{L_{\text{外循环}} \left(\underbrace{L_{\text{内循环}}(\theta; D_{\text{支撑集}}), D_{\text{查询集}}}_{\text{内循环}} \right)}^{\text{外循环}}, \quad (2)$$

其中， $L_{\text{内循环}}(\theta, D_{\text{支撑集}})$ 和 $L_{\text{外循环}}(\theta, D_{\text{查询集}})$ 分别表示在支撑集 $D_{\text{支撑集}}$ 和查询集 $D_{\text{查询集}}$ 上计算的损失。外循环优化的初始模型参数 $\theta^* = \arg \min_{\theta} L_{\text{内循环}}$ 是内循环更新后的模型参数，优化目标转化成学习模型在何种初始条件下可以最小化支撑集和查询集的损失之和。

内循环优化时，首先利用无监督的域适应损失 L_{uda} 替换公式 (2) 的内循环损失 $L_{\text{内循环}}$ ，然后假设潜在退化域图像无标签，使模型从有标签的源域上迁移到无标签的潜在域。外循环优化时，将带有标签的潜在域作为查询集，通过密集预测损失 L_{pre} 评估内循环优化模型在查询集上的域适应性能。通过上述操作，双层循环的优化目标即转化为学习模型的

初始条件 θ_0 ，这个初始条件 θ_0 可以辅助模型更加有效地从有标注数据迁移到无标注数据。因此，公式（2）的双层循环可以表示为如下形式：

$$\theta_0 = \arg \min_{\theta_0} \overbrace{\sum_{D_S, D_L, \overline{D}_L} L_{pre} \left(L_{uda} \left(\theta_0; D_S, \overline{D}_L \right), D_L \right)}^{\text{外循环}} \quad (3)$$

内循环

其中， \overline{D}_L 和 D_L 分别表示实际有标签和假设无标签的潜在域数据。

元测试阶段。通过元训练，可以学习得到的模型的初始参数 θ_0 ，这个参数将被当作元测试阶段域适应模型的初始条件，进一步指导模型从潜在域迁移到目标退化域。当整个训练结束，可以得到收敛后的密集预测网络。在实际应用场景中，通过元学习优化的网络将被用于在真实场景退化域提供可以信赖的密集预测结果。

围绕模型优化，仍存在很多问题需要探索：（1）将进一步考虑不同类型退化域间的关联性，搭建相关性更强的多退化域配对框架，将有利于挖掘出对提高泛化性更有效的先验知识；（2）引入自监督学习策略优化模型，可以建立不同退化域上正交矩空间和语义信息更稳定的映射关系。本项目将深入探索如何优化密集预测模型，增强模型同时到多种复杂场景的泛化能力，以此提高模型对不同类型退化图像的鲁棒性。

4. 特色与创新之处（限 1000 字）

受到环境因素和成像因素影响，自动驾驶场景中图像质量极易发生退化，但是，当前的密集预测方法却普遍存在鲁棒性较差的问题。因此，无法保障始终提供高精度的密集预测结果，影响了密集预测技术的安全性和可靠性，严重降低了该技术的应用价值。面对上述问题，本项目将重点研究基于鲁棒视觉特征的密集预测方法，本项目的**特色之处**在于有效结合正交矩的稳定性和深度学习极强的学习能力，充分发挥二者的协同优势，以“映射空间”为出发点，以“网络架构”为拓展点，以“模型优化”为突破点，最终构建出一套自动驾驶场景中鲁棒密集预测的完备体系，为推动以密集预测为代表的人工智能技术产业化落地打下坚实的基础。本项目的**创新之处**包括以下三点：

（1）**映射空间角度**。当图像质量发生退化时，像素值将发生剧烈变化，破坏了模型映射关系的稳定性，致使无法进行准确的密集预测。为解决该问题，本项目创新性地提出一种基于新型映射空间转换的深度学习密集预测方法，将图像像素空间转换到稳定的正交矩空间，保证模型训练和测试阶段映射关系的高度稳定性，使模型在复杂的自动驾驶场景中，始终可以提供可信赖的高精度密集预测结果。

（2）**网络架构角度**。正交矩空间具有良好的鲁棒性，但是当图像质量发生严重退化时，由正交矩构建的映射关系在高频区域上的稳定性也会受到破坏，降低重要高频区域的密集预测精度。为解决该问题，本项目创新性地提出一种正交矩空间的高频感知网络，通过多任务学习渐进式地实现高频特征感知和密集预测，并在高频特征学习过程加入更强的约束，加强正交矩高频区域密集预测的鲁棒性。

(3) 模型优化角度。域适应方法通过拉近特征空间分布提高模型的泛化能力，但无法保证模型很好地迁移到训练阶段未接触类型的退化域。为了解决该问题，本项目创新性地提出了一种正交矩空间的元学习优化方法，通过多种类型的退化域数据搭建多配对的域适应框架，利用元优化学习出先验知识，用于辅助模型更好地泛化到不同类型的退化域，提高模型在未知退化域上的密集预测精度。

5. 研究计划及预期成果（限 500 字）

5.1 年度研究计划

2022.08-2023.04：充分调研和整理资料，深入剖析正交矩计算方法中存在的问题，构造稳定快速的正交矩计算框架，发挥正交矩和深度学习的协同优势，通过正交矩空间转换实现基于新型映射关系的深度学习方法。本阶段预期发表“三高”论文2-3篇。

2023.05-2023.12：归纳梳理多任务学习方法，剖析深度神经网络结构的设计原理，构造出基于多任务学习的高频感知网络，实现高频区域更稳定的密集预测。本阶段预期发表“三高”论文1-2篇，参加国内外学术会议1-2次，申请国家发明专利1-2项。

2024.01-2024.07：剖析不同类型退化域间的关联性，搭建基于多配对域的优化框架，研究自监督及元学习优化策略，以期在不同退化图像上获取高精度的密集预测结果。本阶段预期发表“三高”论文1-2篇，参加国内外学术会议1-2次，申请国家发明专利1-2项。

5.2 预期成果

(1) 理论成果：首先，通过构建模型更加稳定的空间映射关系，挖掘出鲁棒的视觉特征。然后，通过搭建基于多任务的高频感知神经网络，增强正交矩空间对高频特征感知能力。接着，通过正交矩空间的多域配对元学习优化，提高模型复杂场景下对多目标退化域的泛化能力。最终，形成一套自动驾驶场景下鲁棒密集预测的完备体系。

(2) 学术论文：本项目相关研究成果发表“三高”学术论文4-6 篇，其中力争完成IEEE Trans.汇刊论文 2 篇。

(3) 学术会议：本项目支持参加国内外学术会议2-3 次，获取本研究方向最新动态。

(4) 知识产权：申请国家发明专利2项以上。

(5) 学生培养：协助培养具有一定科研素养的硕士生2-3名，在我校计算机专业形成一个鲁棒视觉表征相关的研究团队。

6. 研究基础（与本项目相关的研究工作积累和已取得的研究工作成绩；已具备的科研条件，尚缺少的科研条件和拟解决的途径；正在承担的与本项目相关的科研项目情况，限 1000 字。）

本人长期从事人工智能领域研究，对计算机视觉技术有比较扎实的研究基础，获得2022年度江苏省人工智能学会优秀博士学位论文奖。丰富的科研项目经验是完成本项目的关键因素，近年，本人作为核心成员参加科研项目的具体情况如下所示：

(1) 基于异质任务耦合的鲁棒视觉行人搜索，国家自然科学基金面上项目，2022/01-2025/12；

(2) 视频中人体检测、再识别和姿态跟踪的联合学习方法，国家自然科学基金中德

国际合作项目，2019/01-2021/12；

(3) 视频中弱小行人目标的快速检测方法，并行与分布处理国防科技重点实验室稳定支持项目，2019/05-2020/04；

(4) 基于深度学习的视频中人体分析技术，CCF-腾讯犀牛鸟科研基金项目，2018/09-2019/09。

截至当前，本人累计发表学术论文23篇，其中，SCI国际期刊16篇，国际会议6篇，国内一级学报1篇。近三年，本人以第一作者身份发表学术论文9篇，其中，5篇收录于中国计算机学会（CCF）推荐的A类和T1类高质量期刊和会议。本项目重点研究基于鲁棒视觉特征的复杂场景密集预测方法，依次从映射空间、网络架构以及模型优化三个层面入手，有效结合正交矩和深度学习方法，旨在构建出一套自动驾驶辅助场景中鲁棒密集预测的完备体系。申请人在近三年的研究工作中，在映射空间，网络架构和模型优化三个方面都进行过深入的研究，积累了比较丰富的理论知识和科研经验，为本项目的顺利实施打下了坚实的基础。申请人前期的具体研究基础如下所示：

1. 正交矩与映射空间方面研究基础。考虑正交矩对图像噪声较强的抵抗能力，本项目将正交矩作为深度学习的输入空间，建立深度学习模型的新型映射关系，提高模型输入和输出间映射关系的稳定性，确保深度学习模型可以生成更加鲁棒的视觉特征，从而实现真实场景下各种退化图像的高精度密集预测。前期研究工作中，申请人针对深度学习的映射空间以及正交矩进行了研究，主要成果及相关研究内容如下所示：

[1] 针对映射空间：提出了一种基于新型映射空间的图像复原方法，将图像空间转换到轮廓波变换域，提升了模型对退化图像高频细节的复原能力；（论文成果列表R5）

[2] 针对映射空间：提出了一种基于剪切波域的图像超分辨率方法，通过建立剪切波域和复原目标空间的映射关系，改善高分辨率图像的重建质量；（论文成果列表R6）

[3] 针对正交矩：提出了一种正交矩建模方法，通过对四元数极谱变换建模，实现数字水印系统鲁棒性和不可见性良好平衡；（论文成果列表R7）

[4] 针对正交矩：提出了一种圆谐傅里叶矩的精确计算方法，有效降低了正交矩计算过程中的积分误差和数值误差。（论文成果列表R8）

[5] 针对正交矩：提出了一种正交矩的特征表示方法，利用正交矩描述图像形状特征，提高了图像检索精度；（论文成果列表R9）

2. 多任务网络架构方面研究基础。考虑高频特征鲁棒性在密集预测系统中的重要性，本项目通过多任务学习搭建高频感知的深度卷积神经网络，结合辅助任务和互信息熵损失函数约束高频特征的表示学习，进一步有效结合正交矩和深度学习模型，提高了重要高频区域密集预测的鲁棒性。前期研究工作中，申请人针对基于多任务学习的网络架构设计进行了研究，主要成果及相关研究内容如下所示：

[1] 针对多任务网络架构设计：提出了一种轻量化图像超分辨率网络，通过先验估计分支为图像重建提供高频细节，提升了图像超分辨率性能；（论文成果列表R1）

[2] 针对多任务网络架构设计：提出了一种基于多任务混合学习的人体语义分割方法，通过多任务学习提高模型在人体图像高频区域的密集预测精度。（论文成果列表R4）

3. 模型优化方面研究基础。考虑优化策略对提高模型泛化能力的重要作用，本项目

利用元学习的双层优化从多域配对域适应框架中学习出元知识，这些元知识可以指导模型更好地迁移到真实自动驾驶场景中多种类型的退化图像，显著提高了模型的鲁棒性和泛化性，确保了深度学习模型在不同退化图像上密集预测的准确性。前期研究工作中，申请人针对模型优化策略进行了深入研究，主要成果及相关研究内容如下所示：

[1] 针对模型优化：提出了一种基于半监督人体语义分割方法，通过抗噪声混合学习优化模型，提高了人工标注数据稀缺条件下模型的密集预测精度；（论文成果列表R3）

[2] 针对模型优化：提出了一种基于元学习域适应的语义分割方法，通过元学习优化，提高模型从虚拟数据到真实自动驾驶场景的泛化能力。（论文成果列表R2）

本项目相关的论文成果列表：

[R1] 本人第一作者. An accurate and lightweight method for human body image super-resolution. IEEE Transactions on Image Processing (TIP), 2021, 30(2): 2888-2897. (CCF A类期刊, 中科院一区期刊)

[R2] 本人第一作者. Learning to adapt via latent domains for adaptive semantic segmentation. In: Proceedings of Conference on Neural Information Processing Systems (NeurIPS), 2021. (CCF A类会议)

[R3] 本人第一作者. Hierarchical information passing based noise-tolerant hybrid learning for semi-supervised human parsing. In: Proceedings of AAAI Conference on Artificial Intelligence (AAAI), 2021: 2207-2215. (CCF A类会议)

[R4] 本人第一作者. Hybrid resolution network using edge guided region mutual information loss for human parsing. In: Proceedings of ACM International Conference on Multimedia (ACM MM), 2020: 1670-1678. (CCF A类会议)

[R5] 本人第一作者. 基于级联密集网络的轮廓波变换域图像复原. 软件学报, 2020, 31(12): 3968-3980. (CCF T1类国内高质量科技期刊)

[R6] 本人第一作者. Single image super-resolution via hybrid resolution NSST prediction. Computer Vision and Image Understanding, 2021, 207(6): 103202. (CCF B类期刊)

[R7] 本人第一作者. Color image watermark decoder by modeling quaternion polar harmonic transform with BKF distribution. Signal Processing: Image Communication, 2020, 88(10): 115946. (CCF C类期刊)

[R8] 本人第一作者. Accurate quaternion radial harmonic Fourier moments for color image reconstruction and object recognition. Pattern Analysis and Applications, 2020, 23(4): 1551-1567. (CCF C类期刊)

[R9] 本人（第一作者）. Improving image retrieval by integrating shape and texture features. Multimedia Tools and Applications, 2019, 78(2): 2526-2550. (CCF C类期刊)

本项目依托××大学××学院，课题的主要研究工作在本学院完成，该学院拥有2个一级学科博士学位授权点；2个博士后科研流动站；3个省部级重点实验室；承担多项国家支撑计划、863计划、国家自然科学基金、国防科研、国际合作、省部级及企事业单位委托项目；多项研究成果获省部科技进步奖、发明创造奖及国际发明展览会金奖等，为科技发展及企业科技进步做出了积极贡献。学院装备各种网络设备、软硬件开发平台等一系列先进的仪器设备和仿真软件。此外学校拥有齐全的中英文期刊数据库，完全满足本项目研究过程中文献的需求。以上均为本项目的顺利开展提供了完备的软硬件条件。